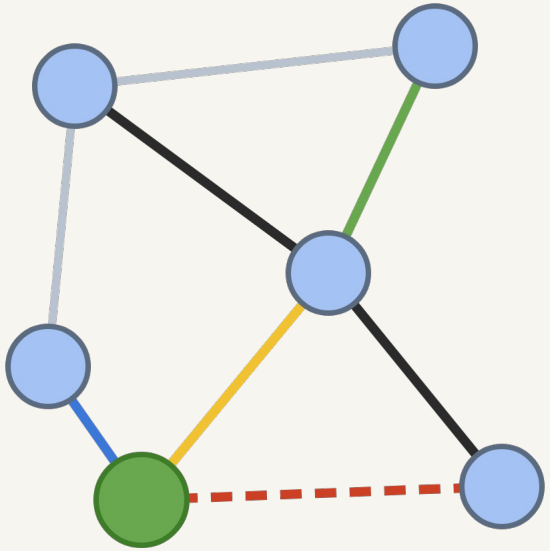
NeurIPS 2026 · Evaluations & Datasets Track

EMERGE

A Benchmark for Updating Knowledge Graphs with Emerging Textual Knowledge

Klim Zaporojets · Daniel Daza · Edoardo Barba · Ira Assent · Roberto Navigli · Paul Groth

Aarhus University · Vrije Universiteit Amsterdam · Eindhoven University of Technology · Amsterdam UMC · Sapienza University of Rome · Principled Intelligence · Babelscape · University of Amsterdam



THE PROBLEM

Knowledge graphs (KGs) store facts as triples. But the world keeps changing.

Growth in the number of Wikidata triples associated with Wikipedia articles

26M → **39M**

January 2019 — January 2025

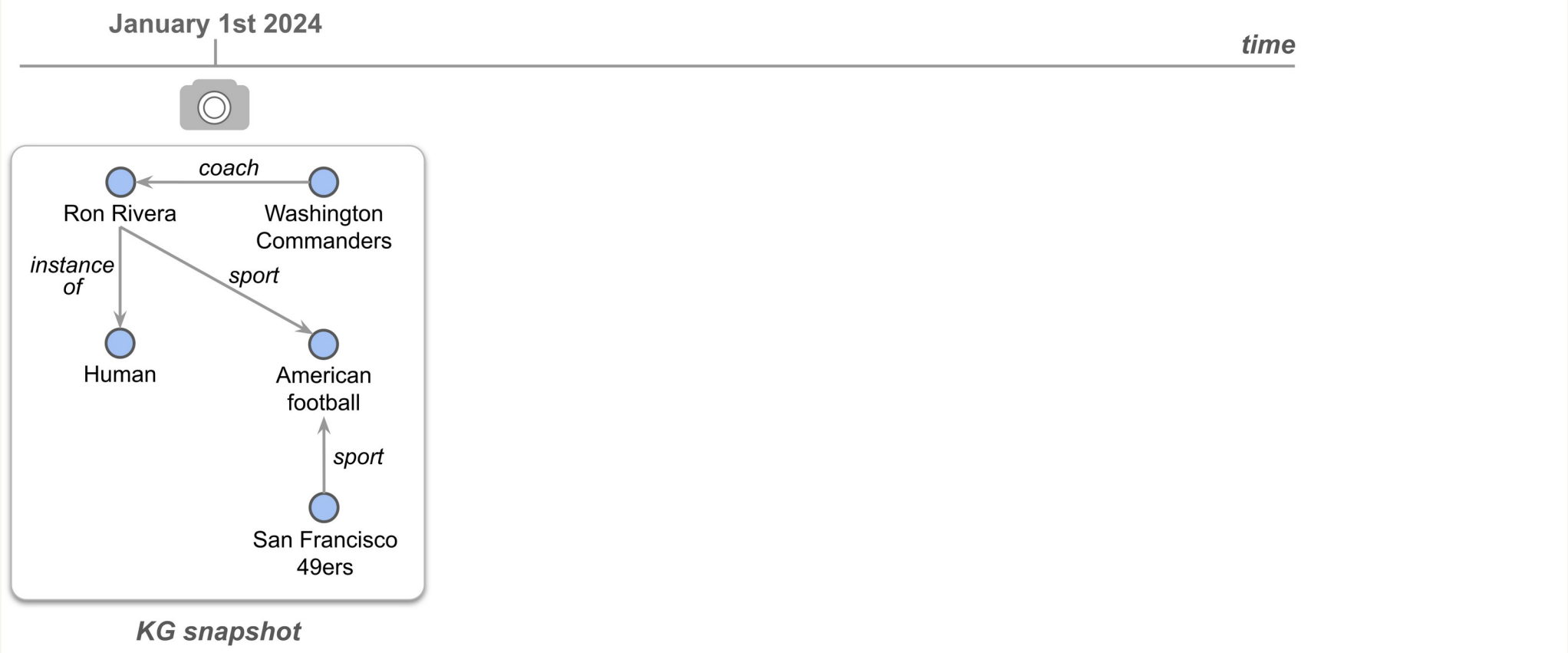
New triples appear every week, and old ones expire.

Can we keep a KG up to date with new knowledge appearing in text?

ONE EXAMPLE · STEP 1

A KG stores facts as triples: (subject, relation, object)

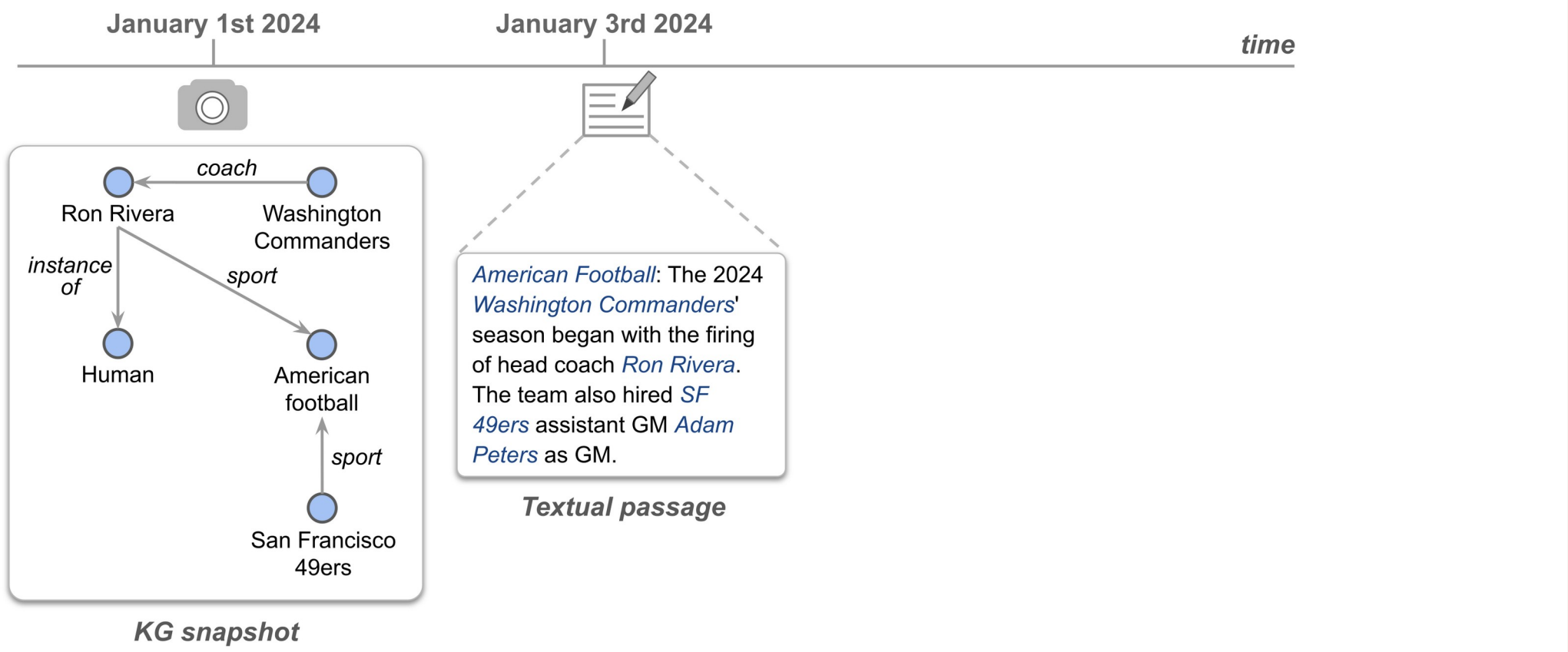
Entities are nodes, relations are edges: (Washington Commanders, coach, Ron Rivera).



ONE EXAMPLE · STEP 2

Two days later, a new Wikipedia passage appears

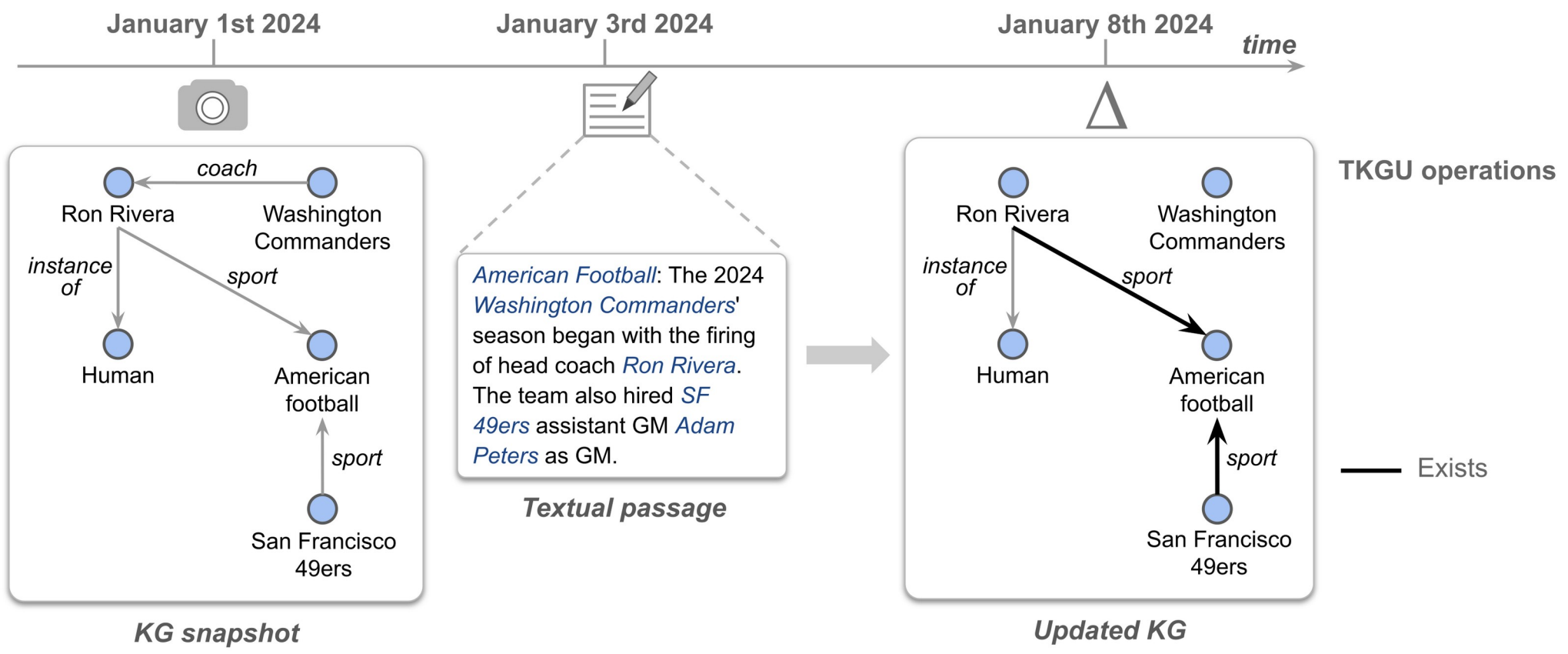
The team fires coach Ron Rivera and hires Adam Peters from the 49ers.



ONE EXAMPLE · STEP 3 · FIVE UPDATE OPERATIONS

1. Triples already in the KG: Exists

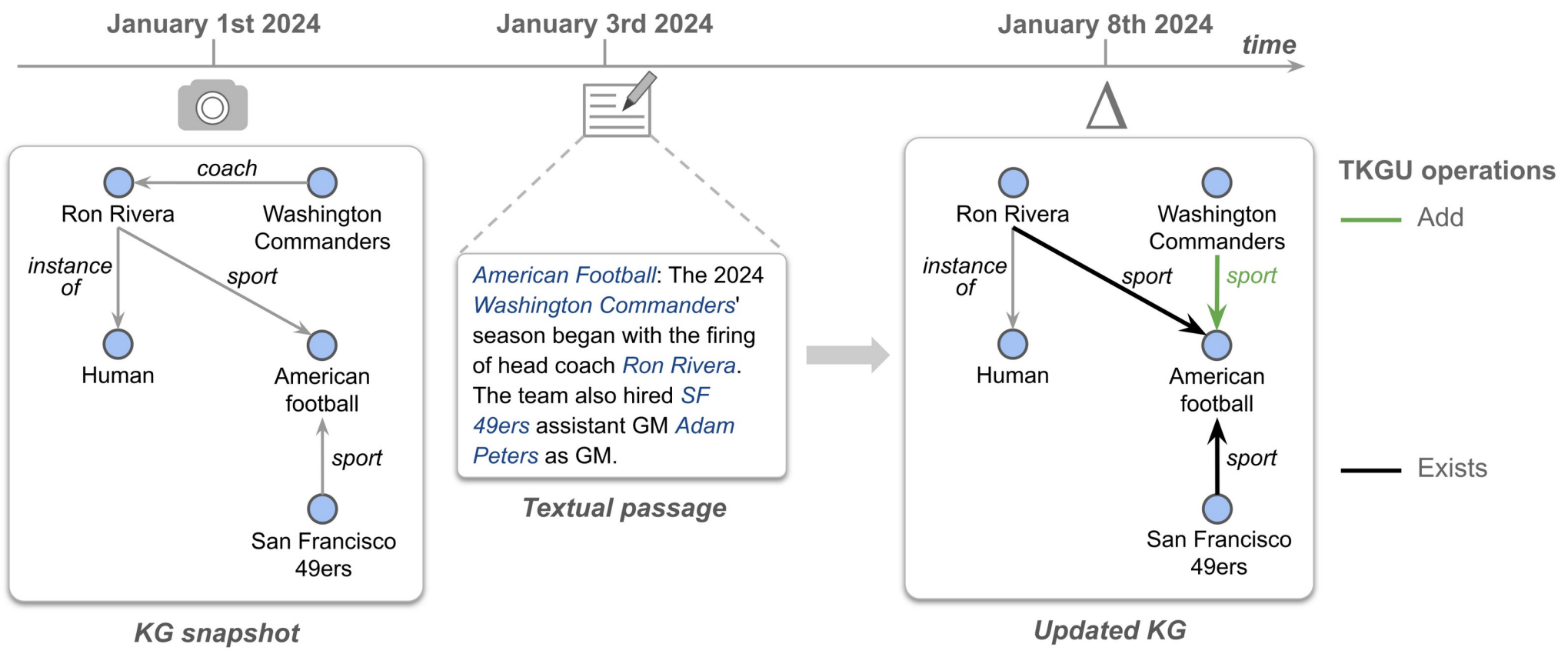
Nothing to change: (SF 49ers, sport, American football) is already there.



ONE EXAMPLE · STEP 4 · FIVE UPDATE OPERATIONS

2. New triples between existing entities: Add

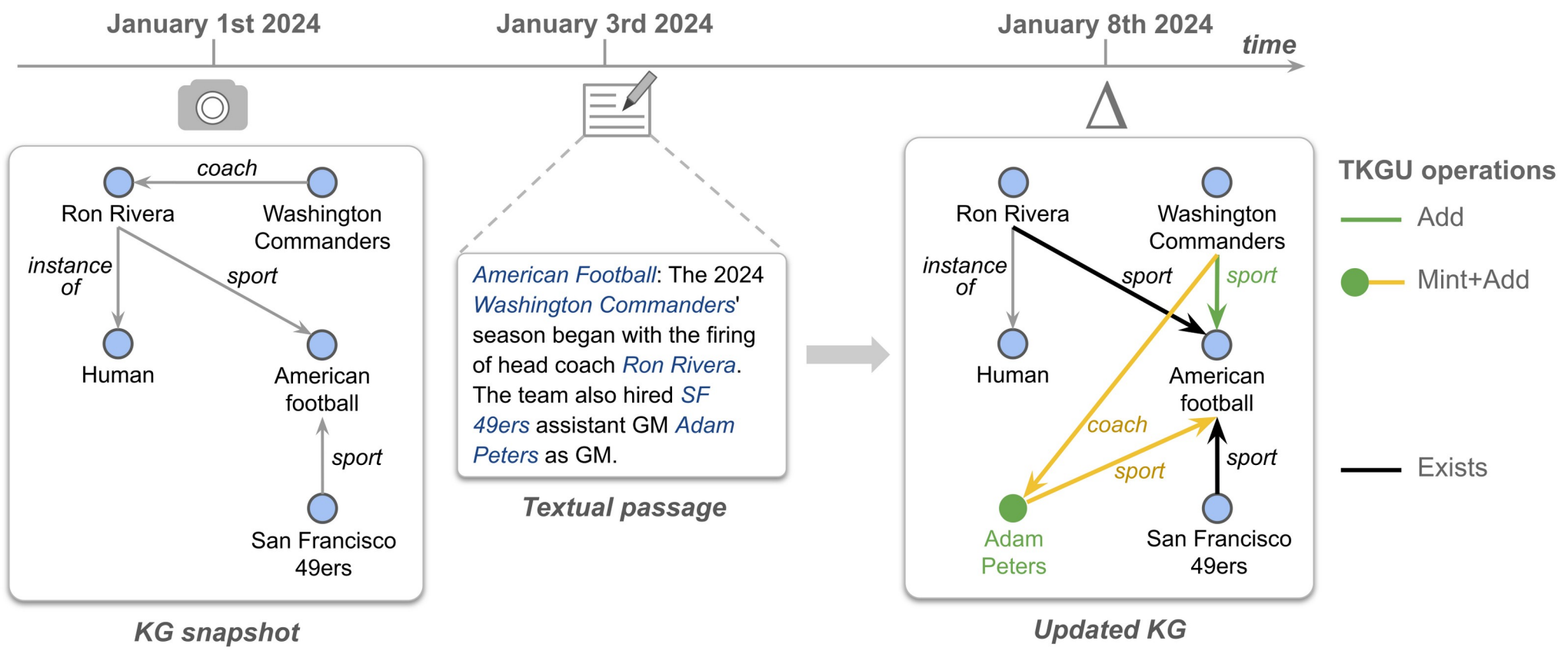
(Washington Commanders, sport, American football)



ONE EXAMPLE · STEP 5 · FIVE UPDATE OPERATIONS

3. A new entity and its triples: Mint+Add

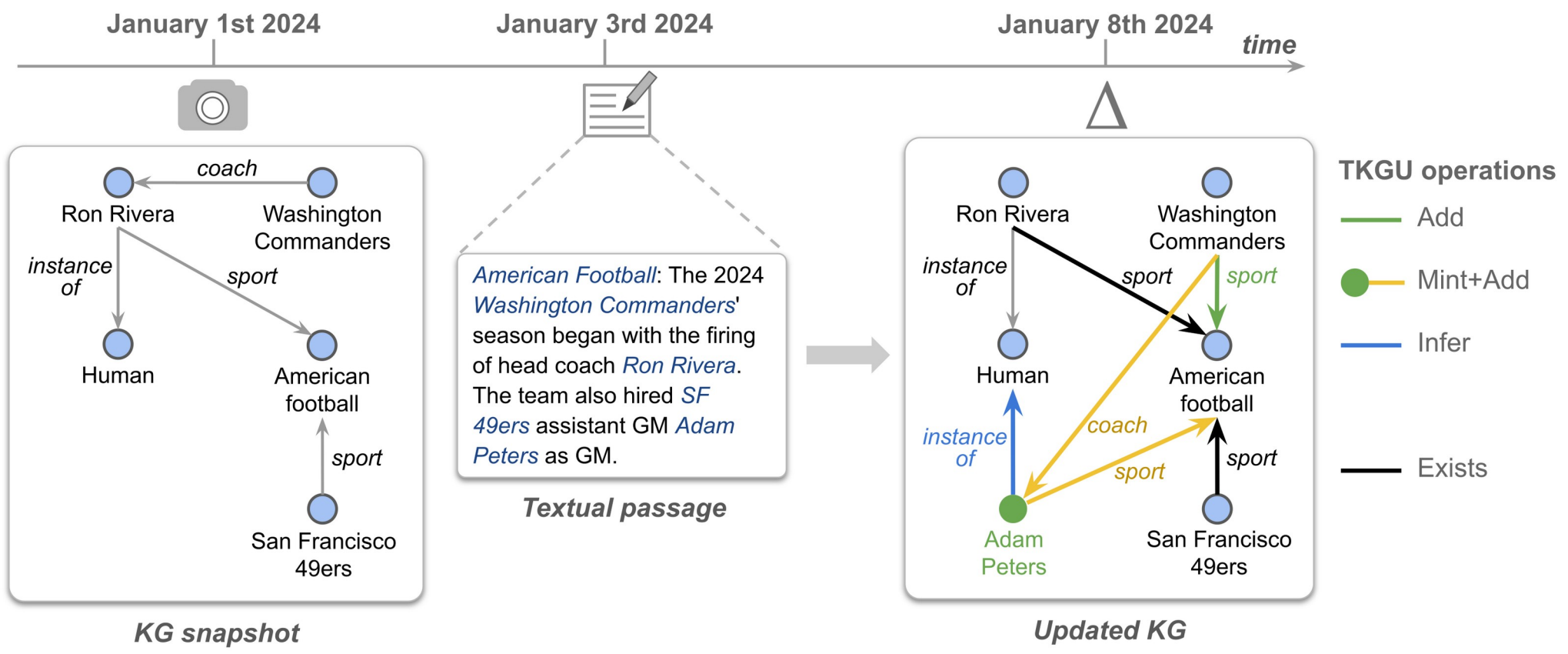
Adam Peters becomes a new node, linked to the team and to American football.



ONE EXAMPLE · STEP 6 · FIVE UPDATE OPERATIONS

4. Triples implied by the KG, not by the text: Infer

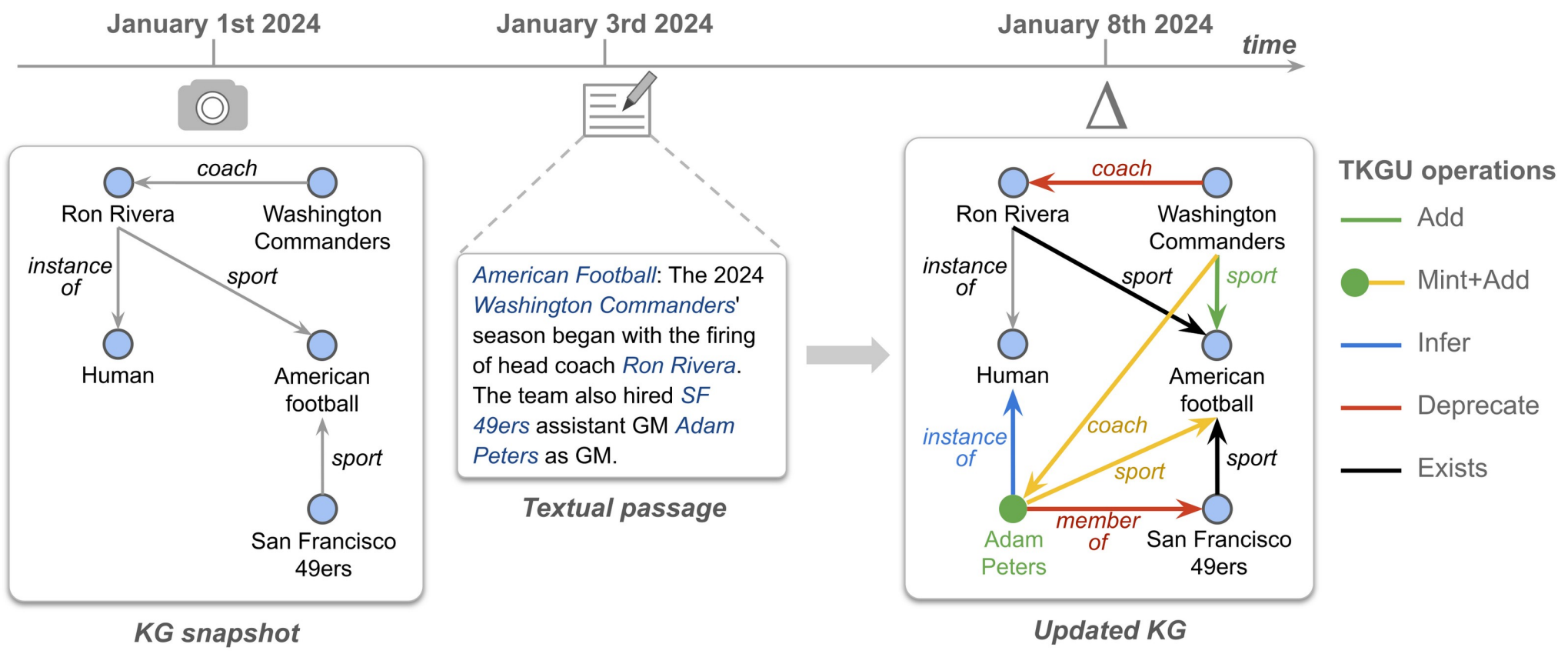
(Adam Peters, instance of, human) is never stated in the passage.



ONE EXAMPLE · STEP 7 · FIVE UPDATE OPERATIONS

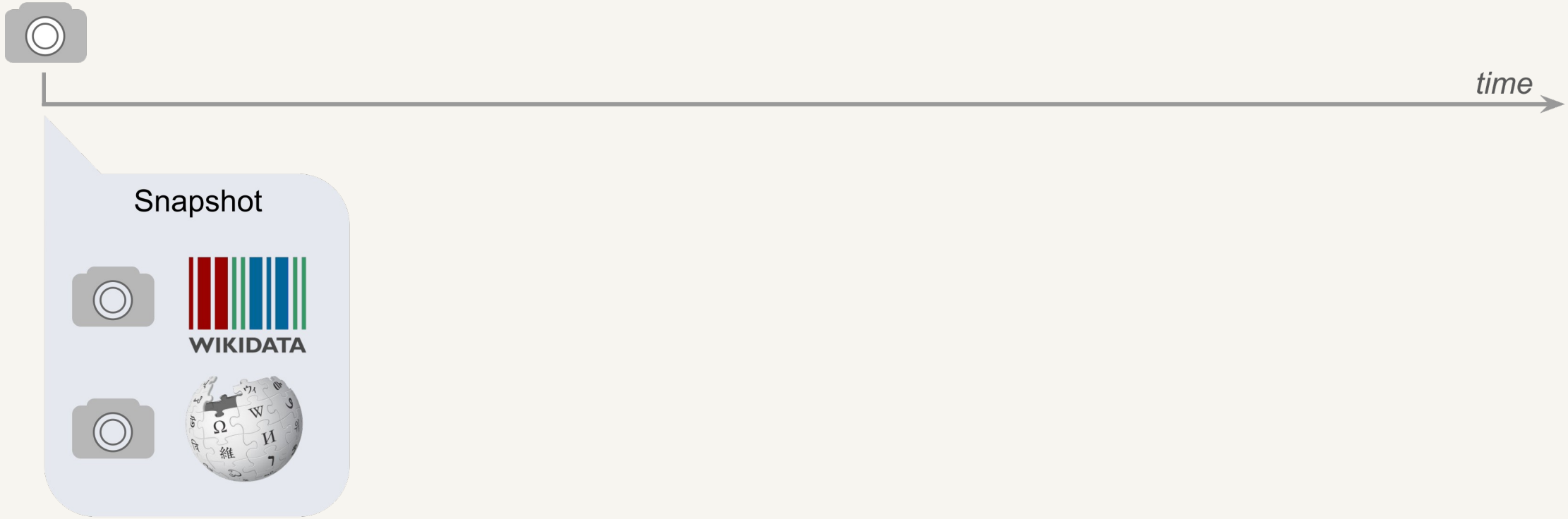
5. Triples that stop being true: Deprecate

(Washington Commanders, coach, Ron Rivera) and (Adam Peters, member of, SF 49ers)



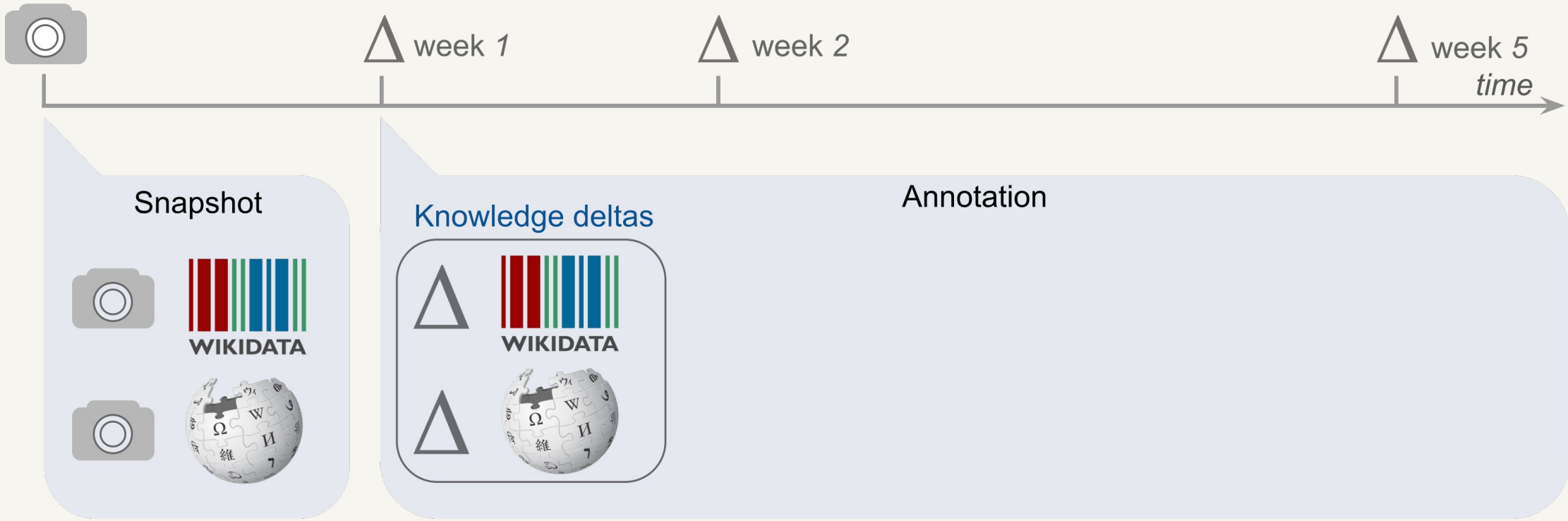
Yearly snapshots: Wikidata + Wikipedia, 2019–2025

Wikidata restricted to entities with a Wikipedia page, as in Wikidata5M.



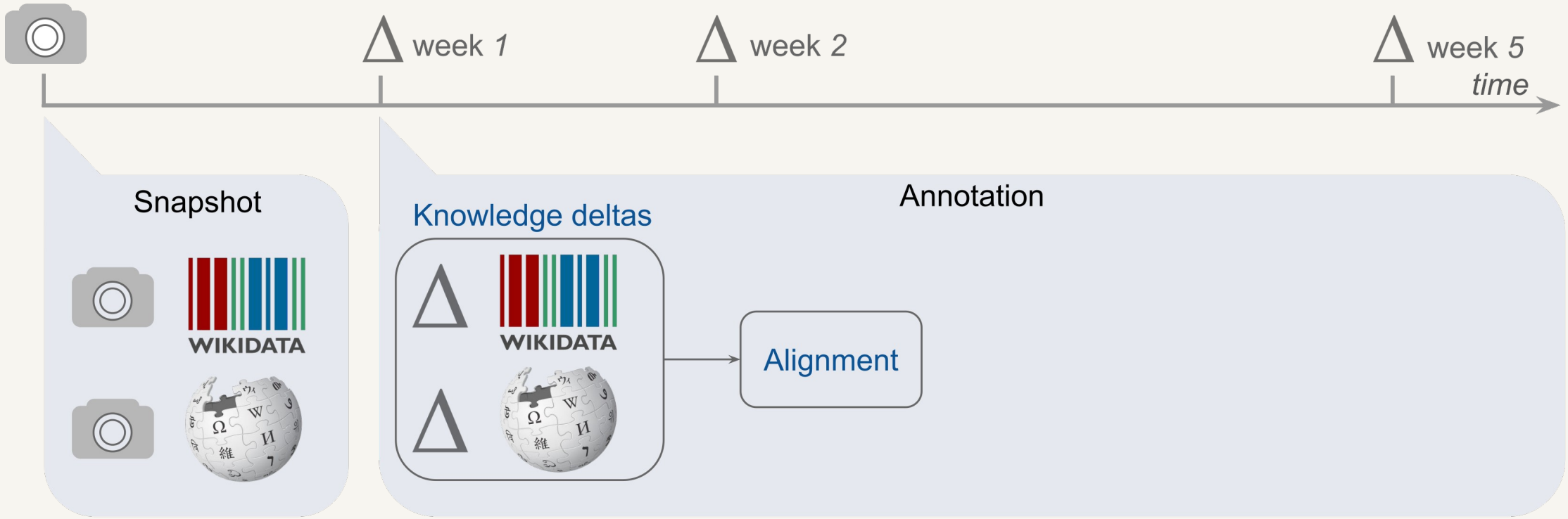
Weekly deltas Δ: changed triples + new passages

Cumulative windows of 1 to 5 weeks after each snapshot.



Alignment: hyperlinked mentions ↔️ Δ triples

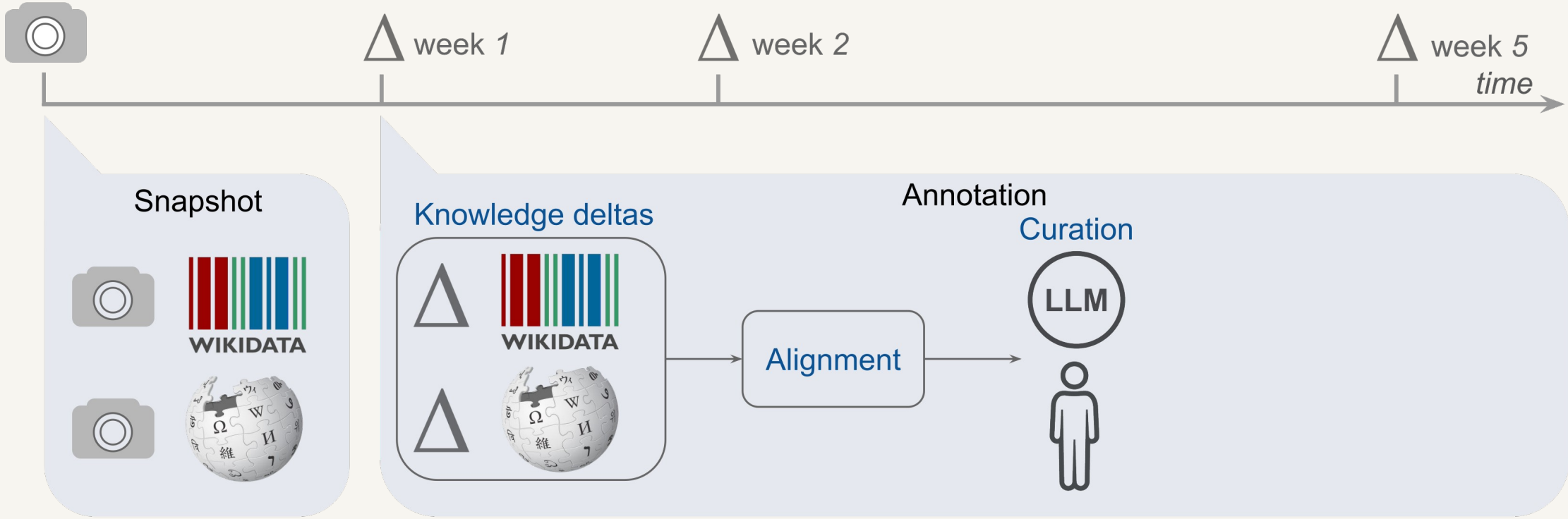
Distant supervision, as in T-REx and REBEL, within the same time window.



EMERGE PIPELINE · 4/5

Curation: an LLM flags unsupported operations

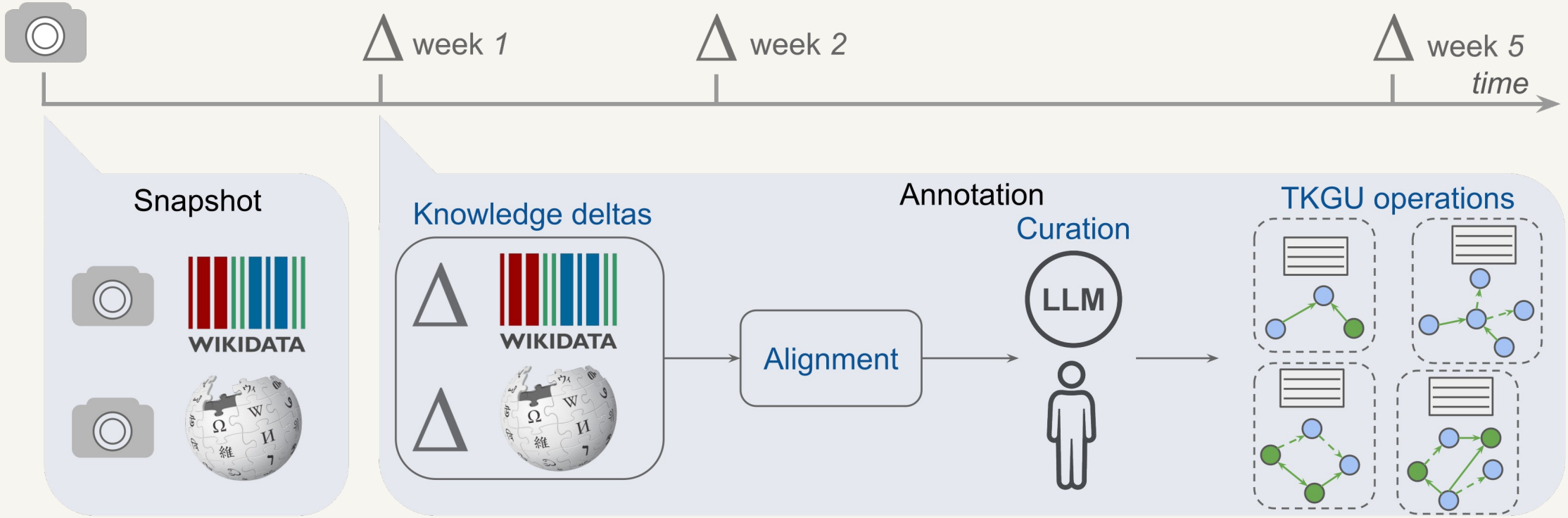
Llama-3.1-405B · agreement with human annotators: Fleiss' κ = 0.77



EMERGE PIPELINE · 5/5

Output: passages paired with their update operations

Exists · Add · Mint+Add · Infer · Deprecate



EMERGE IN NUMBERS

EMERGE at a glance

233K
Wikipedia passages

1.19M
update operations

833
relation types

7
yearly KG snapshots, 2019–2025

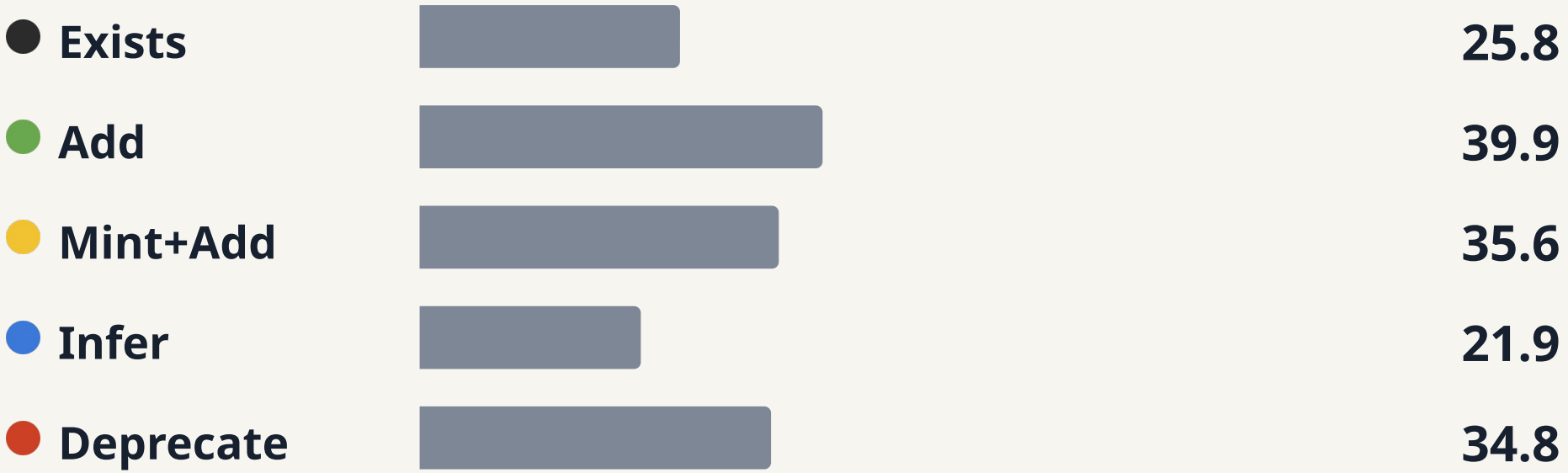
39M
KG triples in 2025 (26M in 2019)

3.5K
test passages, 23.1K update operations

RESULTS

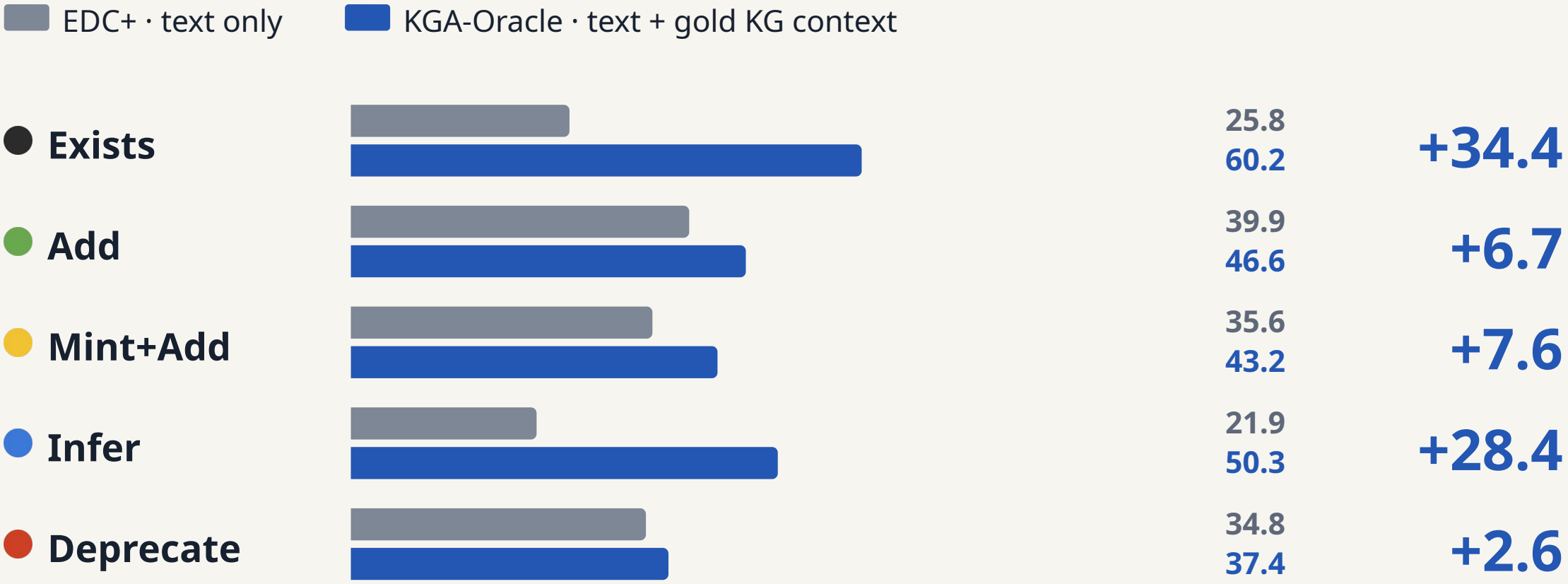
Text-only IE baselines miss most gold triples

Completeness (share of gold triples recovered) of the best IE baseline, **EDC+ with GPT-5.1**



RESULTS

Gold KG context boosts Completeness



RESULTS

KG context makes edits executable, but retrieval is the bottleneck

	EDC+ text only	KGA-RAG KG-RAG: retrieved KG context (top-32)	KGA-Oracle gold KG context
Infer Completeness	21.9	11.3	50.3
Deprecate Executable-R	n/a	12.5	37.6

TAKEAWAYS

EMERGE: a benchmark for keeping KGs up to date with new text

- **All five update operations, from real Wikipedia and Wikidata changes**
- **Text-only IE recovers at most 4 in 10 gold triples**
- **Gold KG context helps a lot, but KG-RAG retrieval is the open problem**

Code
github.com/klimzaporojets/emerge

Data
huggingface.co/datasets/klimzaporojets/emerge-benchmark

Thank you!